A Brief Glimpse at Machine Learning

With Demos and Explanations

Christopher Hittner

11 October 2017

What is Machine Learning?

Machine Learning

- Teaching computers to learn without explicit training.
- Making predictions about data.
    - ex: Predicting movie ratings.
- Learning to perform tasks.
    - ex: Game playing AI for games such as Chess, Go, etc.

Notable AI/ML Applications

- Game Playing AI
  - Deep Blue (1997)
  - AlphaGo (2016)
- Natural Language Processing
  - Cleverbot (1997)
  - Eliza (1966)
- Back-end
  - Google: RankBrain
  - Facebook: Facial Recognition

Learning Problems

- Classification
  - Learning to categorize data.
- Regression
  - Learning to represent continuous functions.
  - Representing more precise probability functions.
- Clustering
  - Categorizing data where the categories are unknown.

Types of Learning

- Supervised Learning
  - Given example inputs and outputs, learn to predict/approximate the data.
- Unsupervised Learning
  - Given only input data, learn to classify it.
- Reinforcement Learning
  - Learner receives stimulus while interacting with environment.

Types of Machine Learning

Classification

- Given a set of items with known categories, classify new items.
  - With a given set of item-category pairs, learn a function to match these pairs.
  - Using classifier, categorize new items into categories.
- Naive Bayes
  - Statistical classification method based on Bayes' Theorem
  - For n features $F = \{F_1, F_2, \ldots, F_n\}$, we seek k such that

$$\underset{k}{\operatorname{argmax}}\, P(C_k|F) = P(C_k) \prod_{i=0}^{n} P(F_i|C_k)$$

  - If C is a continuous set, then P can be replaced with density function p.

Regression

- Given a dataset, we seek a function to fit the data points.
  - Adjust function such that it more accurately represents the data.
- Linear Regression
  - Given a system of linear equations, determine a best fit via a system of equations.
  - To approximate $X\vec{c} = \vec{y}$, we compute $\vec{c}$ that solves $X^T X\vec{c} = X^T\vec{y}$
- Logistic Regression
  - Categorical regression model used to predict probability.
  - We seek to learn parameter $\vec{\theta}$ such that

$$P(x) = \frac{1}{1 + e^{-\vec{\theta}^T\vec{x}}}$$

Clustering

- Given a set of data, group it into categories.
  - We seek to determine which items are alike.
- K-Means Clustering
  - Centroid clustering algorithm.
  - Cluster data into k categories.
  - Points belong to category with closest mean.
- DBSCAN
  - Density-Based Spatial Clustering of Applications with Noise
  - Groups points that are close to one another.
  - Can classify non-linearly separable clusters.

Deep Learning/Neural Networks

Deep Learning

- Form of learning meant to learn features.
  - Generally associated with neural networks.
  - Use of gradient descent to train computation layers.
- Neural Network: Computational model consisting of layers of neurons.
  - Feedforward NN
    - Output computed sequentially.
  - Convolutional NN
    - Computing layers from sections of image.
  - Recurrent NN
    - Layer outputs computed recursively.

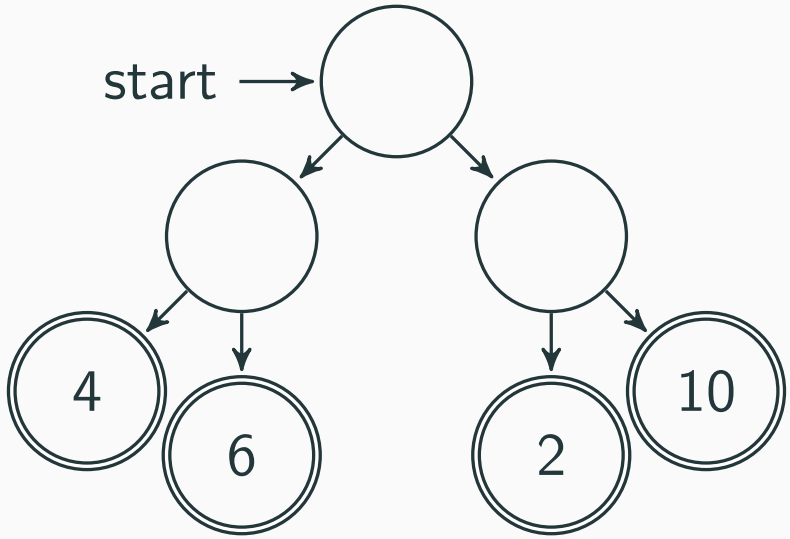
Neural Networks: Learning Algorithms

- Backpropagation
  - Using gradient descent across layers to minimize a loss function.
- Hebbian Learning
  - Increase activation due to given inputs to associate data with stimuli.
- Kohonen Self-Organizing Map
  - Network type used for mapping from high dimensional data to lower dimension.

Decision Trees

Decision Trees

- Given a system with rules, we seek to optimize a score.
  - We want to optimize in long term, not only in short term.
  - Furthermore, we recognize that oppsition can exist.
  - With information on the system, we can create trees of all possible evaluations.



Game: Take turns choosing an edge to traverse. One player maximizes, the other minimizes.

Decision Trees

- Minimax Trees
  - Traverse a tree and find the best possible path.
  - Generally bounded due to large search space.
  - Can be pruned to ignore solutions (Alpha-Beta Pruning)
- Monte Carlo Trees
  - Random search of decision tree to estimate best path.
  - Can prioritize unexplored and favorable solutions during search.
  - Allows deeper search than minimax due to random search.

Natural Language Processing

Natural Language Processing

- Given some spoken or written information, we seek to deduce new information from it.
  - Detecting meaning in a segment of text.
  - Determining relevance of information.
  - Converting between speech and text.
- Word2Vec
  - Algorithm for computing similarity between words.
  - Given a set of sentences, determine representative vectors such that similarity is equal to dot product.
- Lesk Algorithm
  - Predicts approximate definition of word in context.
  - Chooses word definition with most overlap with sentence

Genetic Algorithms

Genetic Algorithms

- We seek to determine optimal parameters in a given scenario.
  - Gradient unavailable; 'fitness' function $F(\vec{\theta})$ has non-trivial gradient.
- Solution: Artificial Selection
  - Evaluate samples and select best ones.
  - Use genetic operators to improve genes (mutation, crossover, selection)
  - Breed an optimal 'gene'.

Further Reading

- *On the Computational Power of Neural Nets* (1995)
- *Neural Network Design* (1996)
- *ImageNet Classification with Deep Convolutional Neural Networks* (2012)
- *Generative Adversarial Networks* (2014)
- *Mastering the game of Go with deep neural networks and tree search* (2016)